UNIONE EUROPEA
Fondo europeo di sviluppo regionale
REPUBBLICA ITALIANA
REGIONE AUTÒNOMA DE SARDIGNA
REGIONE AUTONOMA DELLA SARDEGNA
POR FESR
SARDEGNA 2014-2020
SARDEGNA RICERCHE

Next-Gen Bioinformatics

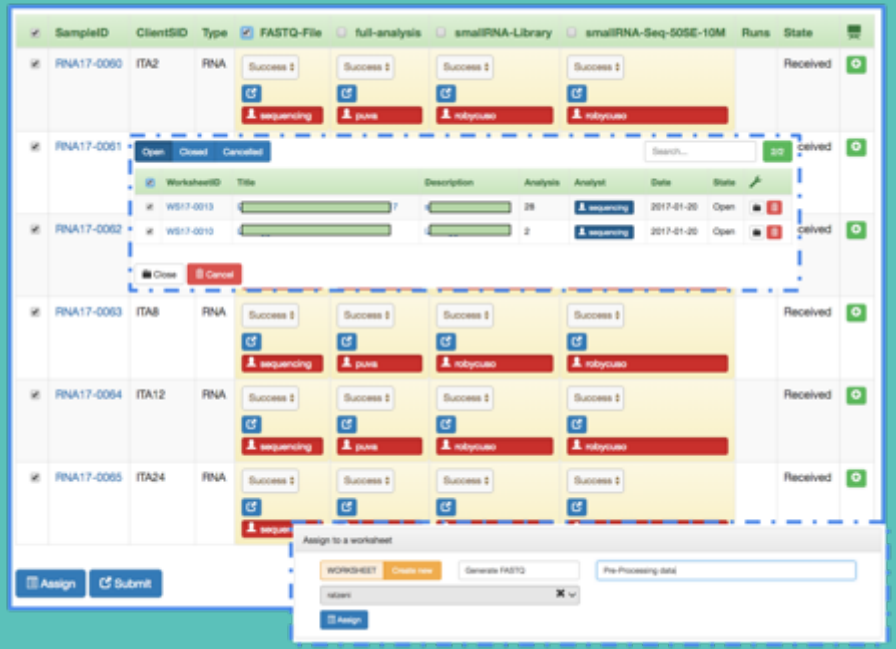
Moving from wet to dry

Gianmauro Cuccuru - gmauro@crs4.it

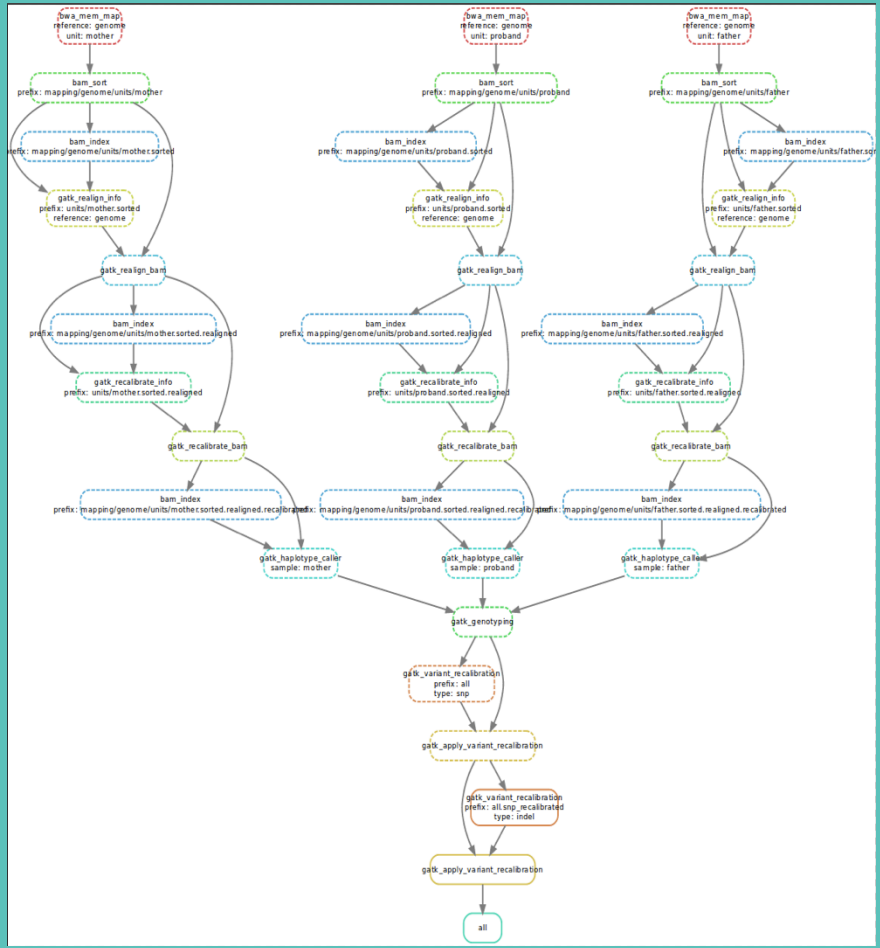
NGS Data Storage requirements

	raw data	pre-processed data
HiSeq 3000	~500 GB	~300 GB
HiSeq 2500	~2,5 TB	~300 GB
HiSeq 2500 rapid	~250 GB	~50 GB

Laboratory Information Management System



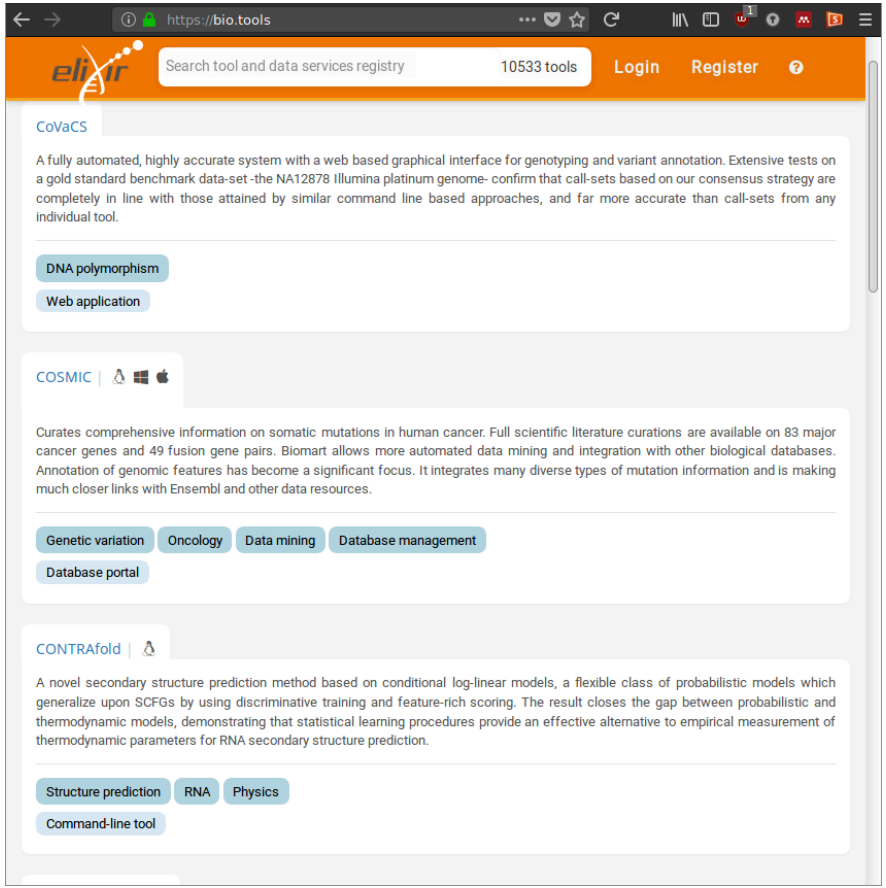
Downstream analysis



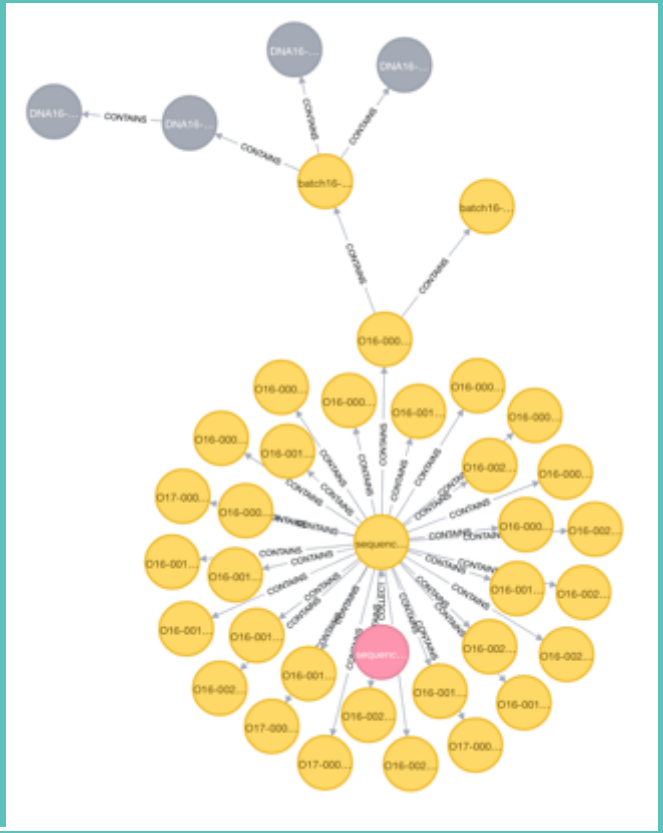
Most tools can:

- run on a single instance
- run on multiple instances
- run against multiples OSes
- be executed from cli
- available only from web
- ...

https://bio.tools



Metadata matters



Enabling computational reproducibility

Documented, portable, comparable, automated software.

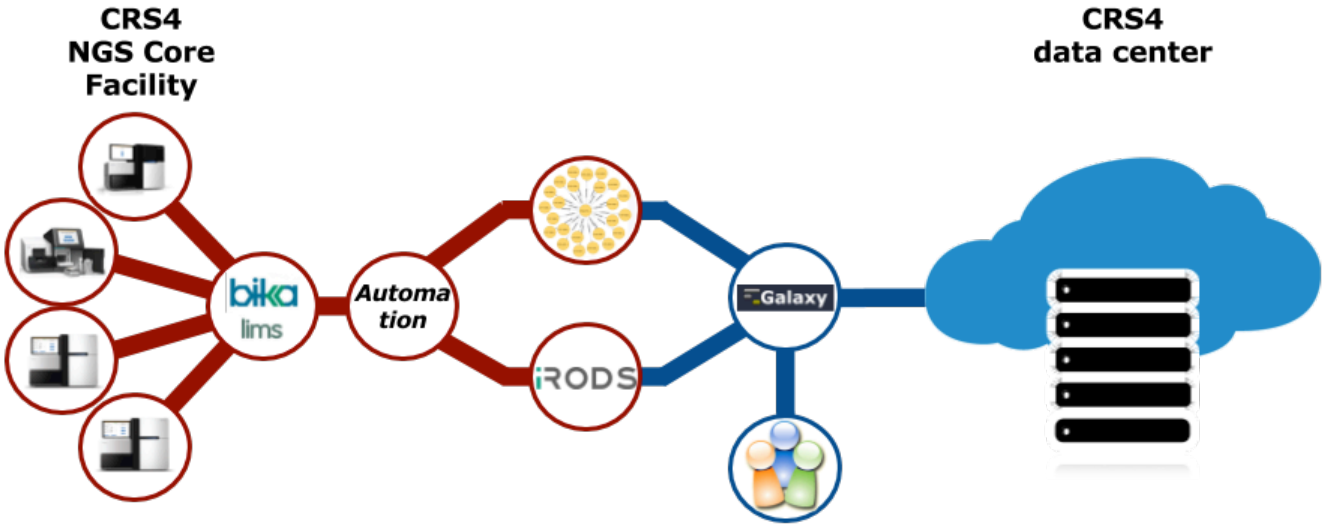
Following this practices:

- develop code under revision control systems
- isolate execution by containers
- record metadata and data provenance
- automated IT steps

Your software should easily run on your own laptop, as well as on HPC cluster, as well as on commercial or research clouds.




Biocomputing Infrastructure @CRS4: Data Management and Computable Framework for NGS studies



DNA sequencing
RNA sequencing
Any NGS application

EXPERIMENT

Data repository
Data integration
Security

ORGANIZE

Genomics
Transcriptomics
Any data analysis workflow

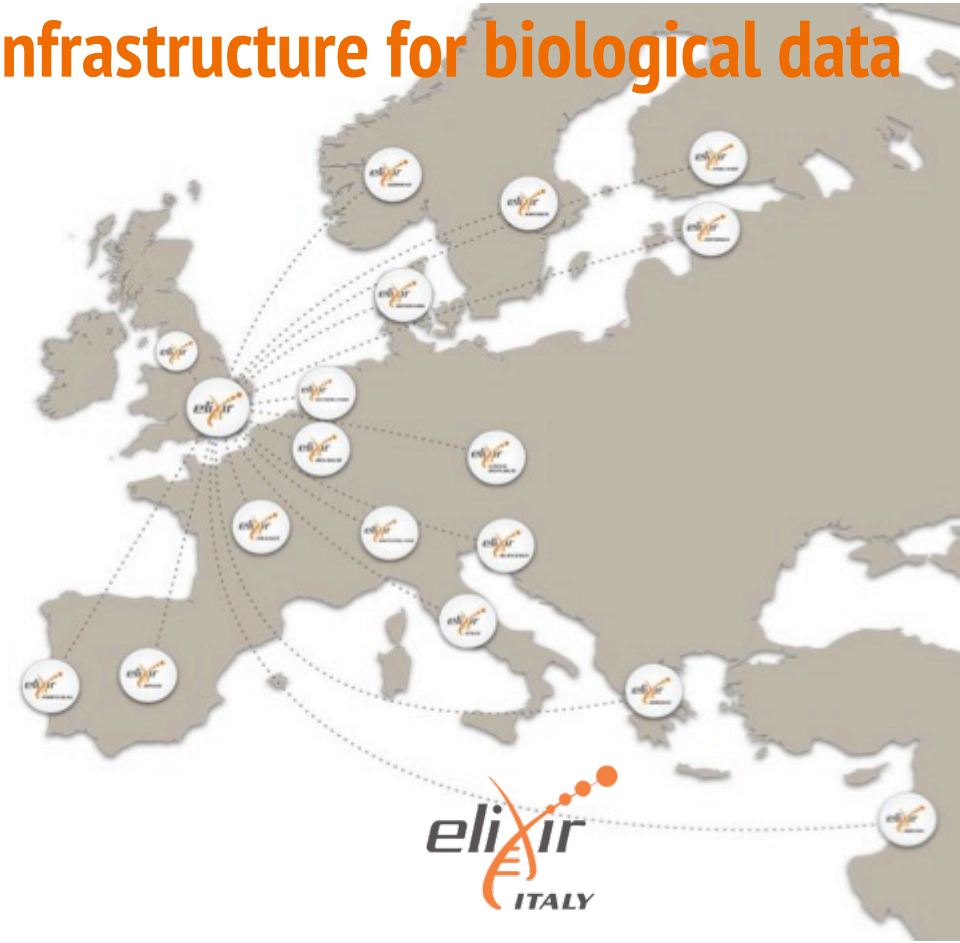
COLLABORATE & ANALYZE DATA

HPC cluster

COMPUTE

ELIXIR: the European Research Infrastructure for biological data

- ELIXIR connects national infrastructures and EMBL-EBI
- 17 Member states + EMBL
  Belgium, Czech Republic, Estonia, Denmark, Finland, France, Israel, Italy, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, Switzerland, UK, Ireland
- ELIXIR deliver services through national ELIXIR Nodes
- ELIXIR Nodes build on national strengths and priorities



http://elixir-italy.org/

Training experience

Galaxy for Bioinformatics tool developers (**ELIXIR IIB Training Programme**)	Development	Cagliari	3-5 July 2017
Exome analysis using Galaxy (**ELIXIR IIB Training Programme**)	Exome analysis	Milano	19-20 September 2016
NGS data analysis with Galaxy	Exome-Seq, RNA-Seq, microbiology	Pula, CA	18-20 November 2014
NGS data analysis with Galaxy	Exome-Seq, RNA-Seq, microbiology	Cagliari	18-20 June 2014 23-25 September 2014
NGS data analysis with Galaxy	Galaxy introduction, QC, microbiology	Bologna	30-31 january 2014
NGS data analysis with Galaxy	Exome-Seq, RNA-Seq, microbiology	Pula, Ca	8-11 June 2013 18-21 October 2013
NGS data analysis with Galaxy	Bacterial sequencing (resequencing & *de novo*), metagenomics	Teramo	12-16 November 2012
Master in Bioinformatics	ChIP-Seq, RNA-Seq, systems biology, protein 3D modelling	Cagliari	6 September, 20 October 2012

100+ researchers from 20+ institutions attended our Galaxy courses.

Biosciences
Data Engineering & Computing

Focusing on:

- integration, interpretation and data intensive analysis of massive and heterogeneous biological data

Our work has the objective of enabling technologically advanced computational and experimental platforms for life sciences and clinical research applications.

http://www.crs4.it/research/biosciences/data-engineering-computing/